

ARI Research Note 92-75

The Precision of Category Versus Continuous Economic Data: Evidence From the Longitudinal Research on Officer Careers Survey

Lucia F. Dunn

Star Network, Inc.

for

Contracting Officer's Representative
Beverly Harris

Leadership and Organizational Change Technical Area
Paul A. Gade, Chief

Manpower and Personnel Research Division
Zita M. Simutis, Director

August 1992



92-25914
SEP 23 1992

92 9 25 065

United States Army
Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

**A Field Operating Agency Under the Jurisdiction
of the Deputy Chief of Staff for Personnel**

EDGAR M. JOHNSON
Technical Director

MICHAEL D. SHALER
COL, AR
Commanding

Research accomplished under contract
for the Department of the Army

Star Network, Inc.

Technical review by

Curtis L. Gilroy
Robert W. Tinney

DTIC QUALITY INSPECTED 3

APPROVED FOR
RELEASE
DATE
BY
A-1

NOTICES

DTIC QUALITY INSPECTED 3

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE			Form Approved GSA FPMR (41 CFR) 101-11.6	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 1992, August	3. REPORT TYPE AND DATES COVERED Final Jul 89 - Jul 90		
4. TITLE AND SUBTITLE The Precision of Category Versus Continuous Economic Data: Evidence from the Longitudinal Research on Officer Careers Survey		5. FUNDING NUMBERS MDA903-88-C-0121 62785A 791 2304 C01		
6. AUTHOR(S) Dunn, Lucia F.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Star Network, Inc. 4480 King Street, Suite 500 Alexandria, VA 22302		8. PERFORMING ORGANIZATION REPORT NUMBER --		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-R 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. SPONSORING/MONITORING AGENCY REPORT NUMBER ARI Research Note 92-75		
11. SUPPLEMENTARY NOTES Contracting Officer's Representative, Beverly Harris.				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.		12b. DISTRIBUTION CODE --		
13. ABSTRACT (Maximum 200 words) This research evaluates category versus numeric responses to questions in the U.S. Army Research Institute for the Behavioral and Social Sciences's Longitudinal Research on Officer Careers (LROC) Survey, which examines career intentions of junior Army officers. The assessment is based on the relative efficiency of estimates of regression model parameters. Efficiency is measured by the standard errors of coefficient estimates of models applied to category and numeric response data, respectively. The analysis consists of two parts. First, a Monte Carlo experiment is conducted. It estimates ordered logit models (OL) for category data and an ordinary least squares (OLS) regression model using numerical response data with measurement error. Second, the analysis of the LROC survey data involves estimation of ordered logit and OLS regression models. The categorical career intentions questions provide data for the dependent variables in the ordered logit models. The dependent variable for the OLS model is the numeric response to the intention question. Sixteen explanatory variables that measure career-related variables (e.g., source of commission and branch satisfaction) and (Continued)				
14. SUBJECT TERMS Junior officers Retention Regression analysis Category variables Continuous variables			15. NUMBER OF PAGES 21	
			16. PRICE CODE --	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

13. ABSTRACT (Continued)

socioeconomic variables (e.g., gender) are included in each model. Findings indicate that standard errors of regression estimates are smaller for numeric than for categorical data. Further, the OLS model applied to the numeric response data confirms a key finding of the LROC project concerning the interaction effect of hours worked per week and branch satisfaction. The categorical response estimates fail to detect this effect.

FOREWORD

The Leadership and Organizational Change Technical Area of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) conducts research to improve the Army's capability to assess manpower and personnel policies by applying empirical analytic techniques and constructing behavioral social science models and data bases. This research note contains empirical data concerning the choice between continuous and categorical survey data from ARI's Longitudinal Research on Officer Careers (LROC) project, which measures Army officer career intentions.

ARI's participation in this effort is part of an ongoing research program designed to enhance the quality of Army personnel. This research is an essential part of the mission of ARI's Manpower and Personnel Research Division to conduct research that improves the Army's capability to effectively manage its personnel and human resource requirements.

THE PRECISION OF CATEGORY VERSUS CONTINUOUS ECONOMIC DATA:
EVIDENCE FROM THE LONGITUDINAL RESEARCH ON OFFICER
CAREERS SURVEY

EXECUTIVE SUMMARY

Requirement:

This research evaluates the accuracy and reliability of continuous response versus category response data about Army officers' career intentions collected as part of the U.S. Army Research Institute for the Behavioral and Social Sciences's (ARI) Longitudinal Research on Officer Careers (LROC) Survey.

Procedure:

The standard errors of regression model parameter estimates are used to evaluate numeric versus category responses to three LROC survey questions on officer career intentions. The analysis consists of two parts. First, a Monte Carlo experiment estimates ordered logit models (OL) applied to category data and an ordinary least squares (OLS) regression model applied to numeric response data with measurement error. The second phase of the analysis involves estimating ordered logit and OLS regression models with LROC survey data. Categorical career intentions questions provide data for the dependent variables in the ordered logit models. The dependent variable for the OLS model is measured by a numeric response intention question. Sixteen explanatory variables that represent career-related factors (e.g., source of commission and branch satisfaction) and socio-economic characteristics (e.g., gender) are included in each model.

The purpose of the Monte Carlo experiment is to isolate the effects of measurement error inherent in numeric response data. This provides valuable information in analyzing the LROC survey data because they include the complex interactions of factors in addition to measurement error.

Findings:

The outcome of the Monte Carlo experiment demonstrates that regression estimates based on continuous, numeric response data are more efficient than estimates with categorical data. This is the case even when the numeric data include large measurement errors and the categorical data have none.

Analyses of the survey data confirm this Monte Carlo result. Most of the OLS coefficient estimates are measured with at least as much precision as the ordered logit estimates. In several cases, the precision of OLS estimates is greater. In particular, the OLS model detects a key interaction effect between branch satisfaction and work hours that is not statistically significant in the ordered logit models.

THE PRECISION OF CATEGORY VERSUS CONTINUOUS ECONOMIC DATA:
EVIDENCE FROM THE LONGITUDINAL RESEARCH ON OFFICER
CAREERS SURVEY

CONTENTS

	Page
INTRODUCTION	1
The LROC Project	1
Categorical Versus Continuous Economic Data: Methodological Issues	2
EMPIRICAL ANALYSIS OF CATEGORY AND NUMERICAL DATA: METHODOLOGY AND RESULTS	3
A Monte Carlo Experiment	3
Evaluation of LROC Survey Data	6
Explanatory Variables	7
DISCUSSION	10
REFERENCES	13
APPENDIX A. THE ANONYMITY OF CATEGORICAL RESPONSE QUESTIONS: THE CASE OF HOUSEHOLD INCOME . . .	A-1
B. EVALUATION OF LROC SURVEY DATA: IMPLICATIONS FOR ORDINAL RANKINGS	B-1

LIST OF TABLES

Table 1. Comparison of Ordered Logit (OL) and OLS Estimates	5
2. Factors Affecting Officers' Intentions to Stay in the Army	9

THE PRECISION OF CATEGORY VERSUS CONTINUOUS ECONOMIC DATA:
EVIDENCE FROM THE LONGITUDINAL RESEARCH ON OFFICER
CAREERS SURVEY

Introduction

This report describes an evaluation of alternative forms of responses to questions in the U.S. Army Research Institute's (ARI) Longitudinal Research on Officer Careers (LROC) Survey about career intentions of company-grade officers. This survey is a key component of a long-term program of empirical research on Army officers conducted by ARI. The objectives of this research program are to identify and track over time the individual, institutional, psychosocial, and family factors and experiences that influence officer career decisions and to specify the policy relevance of these findings for officer manpower, personnel, and force structure planners.

The introduction of this report provides a brief overview of the LROC research program, including the LROC Survey. It also addresses the methodological issue of categorical versus continuous data in empirical research. The methodology and results section describes the research design, data, and results of the empirical analyses undertaken in this research effort. The report concludes with a discussion of implications for the LROC project.

The LROC Project

The LROC research program (a) provides data on the values, attitudes, family situations, and career experiences and plans of the current generation of company-grade Army officers; (b) tests models of the work, career, family, and personal factors that influence officers' decisions to make the Army a career; and (c) provides a rich longitudinal data base for examining the Army experience using behavioral and social science theories and methods. The general strategy for this program involves repeated sampling (over time) of the same subjects. The sample is stratified by commissioning year group and by gender and source of commission within each year group. Data for this research are collected by means of mailed surveys.

Surveys were sent out in 1988, 1989, and 1990 to a stratified random sample of officers (with an overrepresentation of women). Approximately 1000 officers from each commissioning year group (1980 through 1989) were selected for the sample and followed over time. LROC survey items focus on a variety of career and family issues to enable better understanding of the contextual factors that influence the career decisions of officers. Critical to this understanding are survey questions about officer career intentions.

At the present time, there are three questions about career intentions on the survey form. One question asks how many years an officer expects to stay in the Army and provides a continuous response scale from 0-49 years. The other two questions provide broad response categories for length of expected stay. An important methodological question is which way of measuring intentions is "best" for analyses of factors that influence officer career decisions. Answering this question is crucial to an accurate assessment of personnel policy effects on these decisions based on the survey data.

Issues that are relevant to the choice between categorical and continuous response variables are examined in the following section. Empirical evidence concerning which form of the intention data is "best" for analyses of officer career decisions is presented in the results section.

Categorical Versus Continuous Economic Data: Methodological Issues

In conducting empirical policy research, economists and researchers in other social science disciplines focus on model specification issues. These include variables that need to be measured by survey or secondary data sources to estimate parameters and test hypotheses about policy effects. The form the data should be measured in receives relatively less attention however. The alternative ways to measure data items usually found in surveys are: (a) by responses to categorical questions where respondents select a category from a scaled grouping of a continuous variable, and (b) exact numerical responses that directly measure the continuous variable.

Although economic models are often based on continuous response and explanatory variables, both categorical and numerical data are used in empirical economic research. In fact, in recent years, econometric literature has focused on the development of limited dependent (i.e., categorical) variable models (Griliches & Intriligator, 1986; Maddala, 1983). These models address a wide range of economic and empirical issues such as selection bias and unobserved population heterogeneity. Researchers in other social science disciplines tend to rely even more heavily on categorical data.

There are several advantages to using categorical data in empirical research including ease in completing survey forms and coding convenience. Respondents may also prefer to answer categorical questions rather than report exact data.¹ The most important consideration, however, concerns the reliability of reporting by respondents in answering numerical questions (Andrews, 1984; Cox, 1980). Faulty recall by respondents results in measurement error in numerical data. One consequence of this

¹This issue is examined in Appendix A for the reporting of income data.

is that estimates of the effects of policy and other explanatory variables may be biased and inefficient (Judge, Griffith, Hill, Lütekepol, & Lee, 1985). Amemiya (1981), however, demonstrates that estimates based on numerical data are generally more efficient than those derived from categorical data.

Under these circumstances, it is unclear to what extent numerical data are superior to categorical data for empirical research and for the analyses of officer career decisions that are the focus of the LROC project. The methodology and results section below examines the issue of which data source is "best" for analyses of LROC survey data by comparing the efficiency of empirical estimates of parameters obtained for each source.

Empirical Analysis of Category and Numerical Data: Methodology and Results

This section describes a research design that evaluated the alternative questions about officer career intentions on the LROC survey. The results of the evaluation are examined in the discussion.

The LROC survey data reflects a complex of factors in addition to measurement error that may influence officer career intentions, such as unobserved differences between officers in "taste for Army life", random changes in the economy, and unanticipated changes in defense policy (e.g., Operation Desert Storm). In order to provide a "baseline" for the analysis of the LROC survey data, a Monte Carlo experiment was designed in which regression models were estimated using categorical data with no measurement error and numerical data with built-in measurement error, but with no other sources of unobserved variation. The methodology of the Monte Carlo experiment was then applied to the LROC data, and the outcome of the experiment used to interpret the LROC results.

A Monte Carlo Experiment

This section reports the outcome of a Monte Carlo experiment designed to compare estimates of the parameters of an ordered logit (OL) model with ordinary least squares (OLS) estimates of a linear regression model. The logit model estimates are based on categorized data for the dependent variable with no measurement error. The dependent variable for the OLS estimates on the other hand is a continuous variable that includes measurement error.

The OLS model is defined by the following equation

$$Y_i = a + b_1X_{1i} + b_2X_{2i} + b_3X_{1i}X_{2i} + e_{1i}$$

where y_i is the dependent variable, x_{1i} and x_{2i} are exogenous variables, and the error term e_{1i} are independently and identically distributed with mean zero and variance σ_1^2 .

The dependent variable of the logit model is a categorized version of y , defined as

$$Y_{1i} = \sum_{j=1}^J j \cdot 1\{\alpha_{j-1} < y_i \leq \alpha_j\}$$

where $1(A)$ denotes the indicator function for event A , and α_j , $j = 0, \dots, J$ is a set of threshold values where $\alpha_0 = -\infty$ and $\alpha_J = \infty$.

The dependent variable used in the OLS estimation is defined by the equation

$$Y_{2i} = Y_i + e_{2i}$$

where the additional error term e_{2i} , with variance σ_2^2 , represents measurement error or response error. If the error term e_{1i} represents the effect of unobserved heterogeneity between individuals, then there is no response error in the category data in this simulation.

In the Monte Carlo experiment, the normalized standard deviations of the logit estimates are compared with those of the OLS estimates from the underlying continuous variable y without measurement error. The relative efficiency of ordered logit estimation for a given coefficient, denoted by η , is defined as the ratio of the standard deviation of the OLS coefficient estimate to the standard deviation of the logit estimate. The OLS estimator with measurement error is more efficient than the logit estimator (i.e., the observations of y_2 are more informative than those of y_1) provided

$$\sigma_2 < \sigma_1 \sqrt{(1 / \eta^2 - 1)}$$

The exogenous variables x_1 and x_2 are independent standard normal random variables, and the error term e_1 is a logistic random variable with standard deviation $\sigma_1 = \pi/\sqrt{3}$. The latter assumption simplifies the comparison of OLS estimates with conventionally normalized ordered logit estimates. Five categories are defined for the discrete dependent variable y_1 , with equally spaced thresholds $\alpha_j = c \sigma_1 (j - 3/2)$. The regression parameters are $a = 0$ and $b_1 = b_2 = b_3 = b$, where b is chosen such that the variance of the regression function is a specified multiple λ of the error variance σ_1^2 . Different sample designs are characterized by: (a) n , the sample size; (b) c , the category interval width relative to σ_1 ; and (c) λ , the ratio of explained to unexplained variance in the regression.

Table 1 presents results for a sample size of 200, two values of the parameter λ (2 and 4), and three values of the interval width parameter c (0.5, 1, and 2). The normalized standard deviations are the sample standard deviations of the estimates divided by the true values of the corresponding parameters. The magnitude of the measurement error in y_2 is

Table 1

Comparison of Ordered Logit (OL) and OLS Estimates

Category Width (c)	Parameter to be Estimated	Normalized standard deviations of Estimates	Relative efficiency (η)	Equivalent measurement error (w)	
$\lambda = 2.0$					
		OL	OLS		
0.5	b_1	0.151	0.094	0.62	4.3
	b_3	0.165	0.094	0.57	5.0
1.0	b_1	0.126	0.094	0.75	3.1
	b_3	0.123	0.094	0.77	2.9
2.0	b_1	0.148	0.094	0.64	4.2
	b_3	0.144	0.094	0.66	4.0
$\lambda = 4.0$					
0.5	b_1	0.125	0.066	0.53	5.5
	b_3	0.141	0.067	0.47	6.5
1.0	b_1	0.114	0.066	0.58	4.8
	b_3	0.119	0.067	0.56	5.1
2.0	b_1	0.117	0.066	0.57	5.0
	b_3	0.130	0.067	0.51	5.8

determined by assuming that the error term e_2 is uniformly distributed over an interval of width w_0 . The column with heading "equivalent measurement error" reports the maximum values of w for which OLS estimates are at least as efficient as the ordered logit estimates.

Table 1 demonstrates that the precision of least squares estimates based on continuous data with measurement error is greater than logit estimates derived from category data. The standard errors of the OLS estimates are less than those of the OL estimates in each case.

The next section applies the methodology developed here to data from the Army's 1988 Longitudinal Research on Officer Careers (LROC) Survey. The focus of the analysis is on specific properties of continuous and categorical survey data where effects are more complex and difficult to assess than in the Monte Carlo experiment.

Evaluation of LROC Survey Data

Responses to questions in the 1988 LROC Survey about the intentions of officers (i.e., Second and First Lieutenants and Captains) to remain in or exit the Army are analyzed in this section. The survey included three questions designed to capture service intentions. They were as follows:

1. How many years of active duty service do you expect to have completed by the time you leave the Army?

Continuous scale from 0 to 49 years.
2. Which of the following best describes your career intentions?
 - a. I plan to stay in the Army beyond 20 years.
 - b. I plan to stay in the Army until retirement at 20 years.
 - c. I plan to stay in the Army beyond my obligation, but am undecided about staying until retirement.
 - d. I am undecided whether or not I will stay in the Army upon completion of my obligation.
 - e. I will probably leave the Army upon completion of my obligation.
 - f. I will definitely leave the Army upon completion of my obligation.
3. Right now I am (choose one of the following)
 - a. planning on an Army career.
 - b. leaning towards an Army career.
 - c. undecided.
 - d. leaning towards a civilian career.
 - e. planning on a civilian career.

The purpose of the analysis of these questions is to assess the relative performance of category versus numerical data in estimating effects of economic factors on the stay-leave intentions of officers when those data are self reported.² It is important to note that the analysis does not address the related issue of the accuracy of categorical versus numerical response data (Duncan & Hill, 1985; Freeman, 1984; Greenberg & Halsey, 1983; Mellow & Sider, 1983).

Intention to stay in the Army is defined as the dependent variable in this analysis and is related to explanatory variables measured with the 1988 LROC survey data. Three models corresponding to the different intent to stay questions are estimated using the same set of explanatory variables. The dependent variable based on question #1 is analyzed using ordinary least squares analysis. Responses to questions #2 and #3 represent degree of intention to stay in descending order and are dependent variables in ordered logit models. Each of these models is relatively simple in structure and design because the purpose of this research is to investigate the comparative quality of information contained in the three versions of the intentions question rather than to evaluate alternative models of officer intentions.

Explanatory Variables

There are 16 explanatory variables in each model representing job-related characteristics and socioeconomic control variables.³ The variables represent the following factors:

Work-Hours Discrepancy. Hours of work were a major concern for the officers in this sample. Officers in the sample reported working an average of 57.1 hours per week. The mean desired

²It is of interest to assess the self-consistency of responses between the different questions. In order to examine rank-order correlations between the questions, question 1 is divided into the following categories of length of expected active duty service: (a) greater than 20 years, (b) exactly 20 years, and (c) less than 20 years. Using this partition, the following Spearman rank-order correlation coefficients are obtained: (a) question 1 with question 2 -- .81; (b) question 1 with question 3 -- .73; and (c) question 2 with question 3 -- .83. The percentage of missing data for questions 1-3 was 2.2, 0.5, and 0.6 percent, respectively.

³ There are other factors that influence an officer's decision to stay or leave the Army, including the relative attractiveness of civilian alternatives. Analyses of these factors is beyond the scope of this research which concentrates on the factors directly related to Army jobs.

weekly work hours was 47.4. The difference between actual and desired hours of work for each officer is an explanatory variable in the analysis.

Branch Satisfaction. Branch is a major job characteristic for company grade Army officers. There are 24 branches such as Infantry, Aviation, Corps of Engineers, Armor, Military Intelligence, Finance Corps, etc., with widely varying work opportunities. Although officers may request a branch assignment, staffing needs may override the individual's preference. Thirty-two percent of the officers in the LROC sample reported their desired branch to be different from the one to which they were currently assigned. A dichotomous variable (i.e., dummy variable) is included in the analysis to capture a branch effect. The variable is defined as one (1) when the officer is in his or her desired branch and zero (0) otherwise.

Interaction of Work Hours and Branch. This effect is captured by the interaction of work-hours discrepancy and the actual versus desired branch dummy variable defined above.

Source of Commission. The sample was drawn from Reserve Officers' Training Corp (ROTC) and the U.S. Military Academy (USMA) sources of commission. A source of commission variable is defined as one for USMA and zero for ROTC.

Commissioning Year Group. The sample consists of officers commissioned in the years from 1980 to 1987. Commissioning year group is the major determinant of rank, duties, and pay (which is invariant within a rank) because promotions occur at predetermined intervals. Preliminary results indicate that there are no statistically significant differences in intended service for officers commissioned in any year from 1980 through 1984. Officers in these year groups were aggregated into a single category and designated as the reference group for four year group dummy variables.

Spouse Satisfaction. Three dichotomous variables were defined indicating whether an officer had: (a) a spouse dissatisfied with Army life, (b) a working spouse, or (c) a military spouse.

Socioeconomic Variables. Variables representing age, gender, racial/ethnic background, marital status and number of children are also included as explanatory in each model.

The results of the empirical analysis of officer intentions are displayed in Table 2.

The estimates in Table 2 indicate that except for the branch/work-hours interaction term, the coefficient estimates of the three models have the same sign (i.e., positive or negative effect of a given explanatory variable) and are all statistically

Table 2

Factors Affecting Officers' Intentions to Stay in the Army

Estimated Coefficients					
<u>Variables</u>	Numerical Response (OLS)		Interval Scale Categories (OL)		Ordered Scale Categories (OL)
Intercept	20.53 ** (0.43)		0.64 ** (0.11)		-0.23 * (0.12)
Intercept 2	--		-0.98 ** (0.12)		-1.29 ** (0.12)
Intercept 3	--		-2.29 ** (0.12)		-2.17 ** (0.12)
Intercept 4	--		-3.11 ** (0.13)		-3.32 ** (0.13)
Intercept 5	--		-3.96 ** (0.13)		--
Hrs Discr	-0.068 ** (0.021)		0.016 ** (0.005)		0.020 ** (0.006)
Branch	1.477 ** (0.357)		-0.44 ** (0.09)		-0.55 ** (0.10)
HrsxBranch	0.056 ** (0.026)		-0.0057 (0.0067)		-0.0063 (0.007)
Source of Commission	-1.022 ** (0.27)		0.42 ** (0.07)		0.36 ** (0.07)
Yr Grp: 1987	-3.46 ** (0.53)		0.76 ** (0.14)		0.38 ** (0.14)
1986	-3.73 ** (0.47)		0.96 ** (0.12)		0.76 ** (0.13)
1985	-3.33 ** (0.43)		0.89 ** (0.11)		0.73 ** (0.12)
1984	-2.02 ** (0.40)		0.60 ** (0.11)		0.53 ** (0.11)
Spouse Satisfaction	-5.85 ** (0.41)		1.79 ** (0.11)		1.96 ** (0.11)
Working Spouse	-0.91 ** (0.35)		0.29 ** (0.09)		0.23 ** (0.10)
Military Spouse	0.57 (0.40)		-0.027 (0.10)		-0.089 (0.11)
Age	0.44 ** (0.063)		-0.11 ** (0.02)		-0.12 ** (0.01)
Gender	-3.14 ** (0.32)		0.49 ** (0.08)		0.39 ** (0.09)
Race/ Ethnic	0.60 * (0.32)		-0.23 ** (0.08)		-0.22 ** (0.09)
Marital Status	-2.47 ** (0.39)		0.68 ** (0.10)		0.75 ** (0.11)
Number of Children	-0.18 (0.17)		-0.03 (0.05)		0.0056 (0.05)

** 5 percent significance level

* 10 percent significance level

significant at the 10 percent level. Of the 13 explanatory variables where the significance level is at least at the 10 percent level in all three models, the t-ratio using question #1 is higher than with question #2 in 6 cases, and higher than question #3 in 8 cases. In most of the other cases, the differences are small.

The instance where there is a difference in the estimates of the three models is for the effect of the interaction of branch satisfaction and the difference between desired and actual hours of work. Table 2 clearly demonstrates the superiority of the estimate of this interaction effect derived from numerical data. Neither ordered logit model detects this effect using the categorical intentions questions.

The policy implication of this difference is significant. The OLS estimates indicate that the separate main effects of branch dissatisfaction and excessive work hours together result in a loss of 22,400 officer-years from the company grade officer corps. However, the interaction effect between these two variables reduces the main effect of a work-hours discrepancy for officers who are in their desired branch. The impact of an effective work-hours floor is less for officers in their desired branch. Accounting for this interaction effect, the loss to the Army is only 9,140 officer-years, or about 13,000 fewer officer-years than reported above using main effects only.⁴ The inability of the category variable models to detect this interaction results in a substantial over-estimate of the effect of the work-hours discrepancy.

The estimates derived from continuous numerical response data also have the added advantage of being easier to interpret than the categorical response models. The OLS estimates enable analysts to determine the magnitude of the impact of policy relevant variables rather than merely the probability of impact provided by the ordered logit model.

Discussion

This research addresses the issue of the relative precision of statistical estimates obtained from categorical and numerical response questions about officer career intentions in the LROC survey. A Monte Carlo experiment was conducted first in order to simplify the comparison of the different types of data and to isolate the role of measurement error inherent in numerical survey data. The results of this analysis show that ordinary

⁴This estimate is based on the following assumptions: (a) there are 32000 company grade officers in the officer corps, (b) thirty-two percent of company grade officers (the sample average) are not in their desired branch, and (c) the discrepancy between desired and actual hours worked per week is 9.7 hours for all officers and 9.4 hours for officers in their desired branch.

least squares estimates derived from numerical data that include substantial measurement error are more precise than ordered logit models fit to category data with no measurement error.

Secondly, using the Monte Carlo results as a baseline for the purpose of interpretation, analysis of data from the Army's 1988 LROC Survey of company grade officers was undertaken. In this survey, officers were asked to state their military service intentions with one numerical response question and two category questions. The weight of the evidence obtained from this analysis indicates that OLS estimates of the effects of explanatory variables based on the numerical response data are more precise than logit estimates using the categorical data. In addition, the OLS model captures a key interaction effect between branch satisfaction and work hours that is not detected by models estimated with the categorical responses. These results offer credible evidence that numerical data provide more precise and reliable information to examine the policy effects in LROC research.⁵

⁵The results in this section have implications for empirical research in other social sciences as well. See Appendix B.

References

- Amemiya, T. (1981). Qualitative response models: A survey. Journal of Economic Literature, 19, 1483-1536.
- Andrews, F. M. (1984). Construct validity and error components of survey measures. Public Opinion Quarterly, 48(2), 409-442.
- Cox, E. P., III. (1980). The number of response alternatives for a scale: A review. Journal of Marketing Research, 17, 407-422.
- Duncan, G. J., & Hill, D. H. (1985). An investigation of the extent and consequences of measurement error in labor-economic survey data. Journal of Labor Economics, 4, 508-532.
- Eckhardt, K. W., & Ermann, M. D. (1977). Social research methods. New York: Random House.
- Florida Bureau of Economic and Business Research. (1984-1987). Florida Economic Survey. Gainesville, FL.
- Freeman, R. B. (1984). Longitudinal analyses of the effect of trade unions. Journal of Labor Economics, 1, 1-26.
- Greenberg, D., & Halsey, H. (1983). Systematic misreporting and effects of income maintenance experiments on work effort: Evidence from the Seattle-Denver experiment. Journal of Labor Economics, 4, 380-407.
- Griliches, G. & Intriligator, M. (Eds.). (1986). Handbook of econometrics (Vol. 3). New York: North Holland.
- Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., Lee, T. (1986). The theory and practice of econometrics (2nd ed.). New York: Wiley & Sons.
- Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 140.
- Maddala, G. S. (1983). Limited dependent and qualitative variables in econometrics (Econometric Society monographs No. 3). Cambridge: Cambridge University Press.
- Mellow, W., & Sider, H. (1983). Accuracy of response in labor market surveys: Evidence and implication. Journal of Labor Economics, 4, 331-344.

APPENDIX A

The Anonymity of Categorical Response Questions: The Case of Household Income

When respondents do not wish to disclose sensitive information, such as income, they may answer category response rather than numeric questions more accurately because the former provide greater anonymity. This issue is examined here by analyzing income data published by the Florida Bureau of Economic and Business Research (1984-1987).

During the period 1984-87, this bureau conducted a telephone survey of 15,450 respondents selected randomly and asked the following questions about household income:

"Can you tell me roughly what your total family yearly income was before taxes in (year)?"

-- If yes: How much?

-- If no: Well then, could you tell me in which of the following income categories your family would fall?

There were seven income categories respondents could choose starting at less than \$10,000 going up to more than \$45,000 (by \$5,000 intervals.)

Forty-four and a half percent of the respondents sampled reported an exact income figure. Another fifteen percent answered the income category question. Some of the categorical responses may be due to the fact that it is relatively easier to select an income category than it is to recall an exact figure. Consequently, fifteen and a half percent should be interpreted as an upper bound of the fraction of respondents who might prefer the anonymity of categorical questions.

APPENDIX B

Evaluation of LROC Survey Data: Implications for Ordinal Rankings

Economists and other social scientists often measure decision related variables with ordinal scales that represent relative rankings rather than continuous numeric magnitudes (i.e., cardinal measures). For example, characteristics of work environments such as degree of risk and job satisfaction are frequently measured according to ordinal rankings. On a more theoretical level, economists assume that decisions are made by comparing the "utility" of alternatives. In this framework, utility is represented on an ordinal scale, and individual decision makers select the option with the highest utility according to an underlying "utility" function.

Likert (1932) developed a method for ordinal ranking respondents attitudes along a continuum which assigns integer values sequentially to each possible response category of a survey question. A "score" is computed by adding up the answers of all respondents to a given question in a survey. The integers assigned to Likert scale categories are frequently used as if they were continuous data under the assumption "that between each category there is about one unit of distance" (Eckhardt & Ermann, 1977). It is possible to test this hypothesis empirically with limited dependent variable models such as the ordered logit models estimated in the analysis of LROC survey data reported in Table 2 in the text. The estimates of some of the intercepts in the table are consistent with the assumption of a one-unit "distance" between response categories. In particular, the difference between intercepts 1 and 2 (i.e. categories 1 and 2) for question #3 is 1.06. A test was conducted using OLS models with the Likert-type responses to questions #2 and #3 as the dependent variables compared to the OLS model using the continuous response to question #1 as the dependent variable. The results were similar, however, the OLS estimates for questions #2 and #3 did not detect the work-hours/branch interaction effect, duplicating the OL model estimates for questions #2 and #3. This test provides support for attributing the results of the analyses to the type of response data not the type of analyses.